A Panel Data Model for Subjective Information on Household Income Growth

Marcel Das and Arthur van Soest *

Tilburg University

Dept. of Econometrics and CentER

August 1996

JEL-classification: C23, C25, D12

Key words: subjective income expectations, household behavior,
panel data.

Abstract

Subjective expectations about future income changes are analyzed, using household panel data. The models used are extensions of existing binary choice panel data models to the case of ordered response. We consider both random and fixed individual effects. The random effects model is estimated by maximum likelihood. The fixed effects model is estimated by combining conditional fixed effects logit estimates using minimum distance. We find that income change expectations strongly depend on realized income changes in the past: those whose income fell, are more pessimistic than others, while those whose income rose are more optimistic. Expected income changes are also significantly affected by employment status and family composition. Using the same type of models, subjective expectations are then confronted with the head of household's *ex post* perception of the realized income change for the same period. The main finding is that households whose income has decreased in the past underestimate their future income growth.

---

*We are grateful to Rob Alessie, Jeff Dominitz, Ben v.d. Genugten, Peter Kooreman, and Bertrand Melenberg for valuable comments. Data were provided by Statistics Netherlands. Financial support by the Royal Netherlands Academy of Arts and Sciences (KNAW) is gratefully acknowledged by the second author.

Correspondence to: Tilburg University, Department of Econometrics and CentER, P.O. Box 90153, 5000 LE Tilburg, The Netherlands (e-mail: DAS@kub.nl)

1 Introduction

In life cycle models of household behavior, future expectations play a central role. Decisions on consumption, savings, portfolio choice, labor supply, etc., not only depend on current variables, but also on the subjective distribution of future income, prices, etc. (see, for example, Deaton, 1992). In empirical studies of life cycle models, direct information on households' future expectations is rarely used. Instead, the standard approach is to infer expectations from panel data on realizations. [1] This leads to the assumption of rational expectations, or to some alternative explicit model of expectation formation. [2]

Exceptions to this approach are Guiso et al. (1992, 1996), Lusardi (1993), and Alessie and Lusardi (1996), who use characteristics of subjective income distributions directly derived from survey data as explanatory variables to explain consumption, savings or portfolio choice. This type of studies has lead to an increasing interest in data on and the modelling of income expectations. Guiso et al. (1992) and Dominitz and Manski (1996) analyze data on subjective income distributions on the basis of a cross-section. Alessie et al. (1996) use panel data and show that expected changes in income are significantly correlated with actual income changes. Das and Van Soest (1996) explain expected income changes from previous income changes. They also analyze differences between income expectations and realizations over the same time period, and find that many people underestimate their future income, particularly those whose income has fallen in the past.

While Das and Van Soest (1996) focus on one panel wave, this paper uses an unbalanced panel of Dutch households for the period 1984 – 1989. In this way, we can analyze the robustness of the results over time. This is particularly important due to the potential presence of macro-economic shocks, which may imply that results are time

[1] See the discussion in Dominitz and Manski (1996) and the references there.

[2] See, for example, Carroll (1994).

specific. Moreover, it allows for the incorporation of fixed household specific effects. To our knowledge, this is the only survey in which information on income expectations for the same households are available for a number of consecutive years. We focus on income expectations and realizations and use the same survey questions on actual and expected income changes as Alessie et al. (1996) and Das and Van Soest (1996), drawn from the Dutch Socio-Economic Panel (SEP).

The survey questions refer to categories and do not provide information on exact realized or expected income changes. Our dependent variables are therefore of an ordered discrete nature. Although the literature on panel data has expanded rapidly, economic applications of panel data models for discrete data are rather scarce. Examples can be found in Chamberlain (1984) and Pfeiffer and Pohlmeier (1992). [3] Most applications for discrete data consider a binary response. We extend the binary response model to the case of ordered response.

We consider both random and fixed individual effects. The extension in the random effects case is straightforward. In the fixed effects case, we use the conditional logit approach by Chamberlain (1980) after aggregating adjacent categories to two categories. The final estimate for the ordered response model is then obtained by combining the estimates for separate combinations of categories with a minimum distance procedure.

We basically aim at answering the following questions: *Is the use of our type of subjective data feasible and is it useful?* The first question boils down to asking: *do the answers make sense?* We claim that they do, by describing them for the six years and by showing that their relation to various background variables is rather robust over time and of the expected sign. The second question can be restated as: *are the subjective data in conflict with the usual assumptions on rational expectations and (absence of) macro-economic shocks?* Our analysis of the deviations between expectations and realizations suggests

[3] More applications exist in the fields of biology, psychology and biomedicine. An example of the latter is Gibbons et al. (1994).

that they are, and that the assumptions on rational expectations or absence of macro-economic shocks are not valid. This makes it worthwhile to replace these assumptions by information based upon the subjective information in the data.

The organization of the paper is as follows. In section 2 we formulate the panel data model for the ordered responses. In section 3 we use this model to explain income change expectations. Among the explanatory variables are the actual income level and information on the realized income change during the previous year. To see whether different social groups have (ceteris paribus) different income expectations, we also include dummy variables for being unemployed, disabled, or retired. In section 4 we first look at subjective information on realized income changes and show that it relates quite well to more traditional measures of income change, at least on average. We then use the same type of econometric model to compare the expectations in year t with the realizations in year $t+1$ ($t = 1984, \ldots, 1988$). The dependent variable is then based upon the difference between the answers to the questions on expected and realized income changes. Finally, in section 5, we summarize our findings.

2 Panel data models for ordered categorical data

Our starting point is the well-known binary choice panel data model with time varying parameters and individual effects:

$$\begin{aligned} y^*_{i,t} &= \beta'_t x_{i,t} + \alpha_i + u_{i,t}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T \\ y_{i,t} &= 1(y^*_{i,t} \geq 0) \end{aligned} \tag{1}$$

in which $\beta_t \in \mathbb{R}^k$ and $1(A)$ is the indicator function which is equal to 1 if A is true and 0 otherwise. The index i represents the household and index t represents time. Instead of observing $(y^*_{i,t}, x'_{i,t})'$ one observes $(y_{i,t}, x'_{i,t})'$, in which $x_{i,t}$ is a k-dimensional vector of explanatory variables, including a constant term.

We assume that x_i and u_i are independent, where $x_i = [x'_{i,1}, x'_{i,2}, \ldots, x'_{i,T}]'$ and $u_i = [u_{i,1}, u_{i,2}, \ldots, u_{i,T}]'$. The mutually independent disturbances $u_{i,t}$ are assumed to follow some distribution with mean 0 and variance σ^2. In this paper we consider the normal and the logistic distribution.

It is straightforward to extend model (1) to allow for more than two outcomes for $y_{i,t}$. Suppose $y_{i,t}$ can take p possible outcomes. As in model (1), these outcomes are assumed to be determined by an underlying latent variable $y^*_{i,t}$. The relation between $y_{i,t}$ and the underlying latent variable is modelled by

$$\begin{aligned} y^*_{i,t} &= \beta'_t x_{i,t} + \alpha_i + u_{i,t}, \qquad && i = 1, \ldots, N, \quad t = 1, \ldots, T \\ y_{i,t} &= j \quad \text{if } m_{j-1} < y^*_{i,t} \leq m_j && j = 1, \ldots, p \end{aligned} \tag{2}$$

where $m_0 = -\infty$ and $m_p = \infty$. To identify the model, location and scale have to be fixed.

For the individual effect α_i we will discuss two specifications. In section 2.1 the individual effect is assumed to be random and in section 2.2 the individual effect is treated as a fixed effect.

2.1 Random effects specification

The random effects model consists of model (2) together with additional assumptions on the *random* individual effect α_i. We assume that α_i is normally distributed with mean 0 and variance σ^2_α. [4] Moreover, we assume that $x_i, u_i,$ and α_i are independent.

In general, the likelihood function for model (2) is a T-variate integral. However, under the assumption of independence made above, the multivariate integral can be reduced to a single integral by integrating out the individual effect. The integrand is then a product of one normal density and T differences of values of the distribution function F_σ of $u_{i,t}$, (with σ a scale parameter) [see Butler and Moffitt (1982)]. The contribution $\text{Prob}(y_{i,1}, \ldots, y_{i,T})$

[4] For random effects models in which the assumed family of distributions for the individual effect adopts a variety of forms and shapes, see Crouchley (1995).

for individual i to the likelihood function is given by

$$\int_{-\infty}^{\infty} g(\alpha_i)[\prod_{t=1}^{T}\{F_\sigma(m_{y_{i,t}} - \beta_t' x_{i,t} - \alpha_i) - F_\sigma(m_{y_{i,t}-1} - \beta_t' x_{i,t} - \alpha_i)\}]\mathrm{d}\alpha_i, \quad (3)$$

where $g(\alpha_i)$ is the density of $N(0, \sigma_\alpha^2)$. The boundaries $m_j (j = 1, \ldots, p-1)$ are assumed to be constant across individuals.

The model described so far is only applicable for balanced panels. Since the data set we use in our analysis is unbalanced, the notation should slightly be adapted. Define

$$c_{i,t} = \begin{cases} 1 & \text{if individual } i \text{ is in wave } t \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

We assume that $c_{i,t}$ is independent of $u_{i,t}$ and α_i, implying that we do not allow for selection or attrition bias. The likelihood contribution for individual i is then given by [cf. (3)]

$$\int_{-\infty}^{\infty} g(\alpha_i)[\prod_{t=1}^{T}\{F_\sigma(m_{y_{i,t}} - \beta_t' x_{i,t} - \alpha_i) - F_\sigma(m_{y_{i,t}-1} - \beta_t' x_{i,t} - \alpha_i)\}^{c_{i,t}}]\mathrm{d}\alpha_i.$$

2.2 Fixed effects specification

One major limitation of the random effects specification is the assumption that the individual effect α_i is uncorrelated with the $x_{i,t}$. This can be relaxed by treating α_i as a fixed effect implying that each α_i becomes an unknown parameter. In the fixed effects specification, the levels of the slope coefficients $\beta_{t,k}$ are only identified if the corresponding regressors $x_{i,t,k}$ vary over time. For time-invariant $x_{i,t,k}$, only the differences $\beta_{t,k} - \beta_{s,k}$ are identified, implying that without loss of generality, the coefficients of one time period can be normalized to zero.

In this fixed effects model, the number of parameters increases with the number of individuals N. ML estimates of the α_i and the $\beta_{t,k}$ will be inconsistent if N becomes large but T is finite. This is known as the incidental parameter problem (Neyman and Scott, 1948). For the binary choice panel data model, Chamberlain (1980) suggested an approach based upon a conditional likelihood function to estimate the $\beta_{t,k}$. The key idea is to work with a conditional likelihood function, conditioning on sufficient statistics for the nuisance parameters α_i. This idea works if the disturbance terms $u_{i,t}$ are iid and follow a logistic distribution. In that case the minimum sufficient statistic for α_i is $\sum_{t=1}^{T} y_{i,t}$. Given this statistic, the contribution of individual i to the conditional likelihood function is, in case of a balanced panel

$$\text{Prob}(y_{i,1},\ldots,y_{i,T}|\sum_{t=1}^{T} y_{i,t}) = \frac{\exp[\sum_{t=1}^{T}(x'_{i,t}\beta_t)y_{i,t}]}{\sum_{d\in B_i}\exp[\sum_{t=1}^{T}(x'_{i,t}\beta_t)d_t]}, \tag{5}$$

where

$$B_i = \{d = (d_1,\ldots d_T) \mid d_t = 0 \text{ or } 1, \text{ and } \sum_t d_t = \sum_t y_{i,t}\}.$$

It does not depend on the incidental parameters α_i and the conditional ML estimator of β_t is, under mild regularity conditions, consistent and asymptotically normal.

A direct extension of this approach to an ordered response panel data model where the dependent variable has $p > 2$ possible outcomes, is not straightforward and even seems impossible. However, we can combine adjacent categories so that the dependent variable is summarized as a binary variable, and then use the conditional logit method. If we repeat this for all the possible combinations of adjacent categories, we get $p-1$ estimates of the parameters of interest. [5] These estimates can then be combined into one final estimate of the parameters of interest by using minimum distance. See Appendix A for some details.

It is straightforward to extend this estimation procedure to the case of an unbalanced panel. Again, the notation should slightly be adapted. We define $c_{i,t}$ as in (4) and assume

[5] The boundaries m_j are not estimated and can be seen as nuisance parameters.

that $c_{i,t}$ and $u_{i,t}$ are independent to exclude attrition and selectivity bias. Then the conditional probability for the binary case [cf. (5)] is given by

$$\text{Prob}(y_{i,1},\ldots,y_{i,T}|\sum_{t=1}^{T} y_{i,t}) = \frac{\exp[\sum_{t=1}^{T} c_{i,t}(x'_{i,t}\beta_t)y_{i,t}]}{\sum_{d\in B_i}\exp[\sum_{t=1}^{T} c_{i,t}(x'_{i,t}\beta_t)d_t]}.$$

The unbalanced nature of our data is also the reason why we do not consider quasi fixed effects models [see Chamberlain (1984)] in which α_i is allowed to be correlated with the $x_{i,t}$. The fact that $x_{i,t}$ is unobserved in some waves would then lead to ad hoc adjustments of the correlation pattern (or to joint modelling of the $x_{i,t}$ with the $y_{i,t}$ and the specification and computational problems involved with that).

3 Income change expectations: data and estimation results

Data are taken from the Dutch Socio-Economic Panel (SEP), which is a random sample from the Dutch population, excluding those living in special institutions like nursing homes. [6] Households were interviewed in October 1984 and then twice a year (April and October) until 1989. Since 1990 the survey has been conducted only once a year in May. In the October interview, information about income is gathered. We focus on the waves of 1984 till 1989, because in 1990 the questions related to (actual) income have changed substantially.

The attrition rate in the panel is about 25 percent on average, and tends to decrease over time. New households have entered the panel each year. After eliminating observations with item nonresponse, mainly due to missing information on one or more components of actual household income, we retained a sample of 6845 households. Only

[6] See CBS (1991) for details about contents, setup and organization of the SEP.

722 of them are in the balanced subpanel (10.5%). This is the reason why we do not estimate the model for the balanced subpanel only, but focus on the unbalanced panel. For 14% of all households the required information is available in five waves, for 18% in four, for 16.8% in three, and for 16.4% in two waves. The remaining households (24.3%) provided information for only one wave. Most of those who are in more than one wave, participate in consecutive waves. In the final data set used for estimation, about 24% are included in non-consecutive waves, mainly due to item nonresponse. The numbers of observations per wave are included in Table 1.

Heads of households are asked to answer the question

> *What will happen to your household's income in the next twelve months? Possible answers: strong decrease (1); decrease (2); no change (3); increase (4); strong increase (5).*

The distribution of the answers, which will be denoted by EXP_t ($t = 84, \ldots, 89$), are given in Table 1. We see that except for 1984 the number of households expecting a strong decrease is relatively low. If we aggregate the categories *strong decrease* and *decrease* we see that, with the exception of 1987, the number of households expecting a fall in household income decreases. This is also reflected in the mean value of EXP_t: it generally increases, with a small drop in 1987.

Table 1 : Univariate frequencies (in %) of $\mathbf{EXP}_t$ ($t = 84, \ldots, 89$)

EXP_t	84	85	86	87	88	89
# observations	2683	2787	3850	3899	4059	4133
1: strong decrease	5.9	1.9	1.6	2.3	1.3	1.3
2: decrease	33.1	18.9	12.6	15.8	10.9	8.2
3: no change	50.3	62.4	66.4	63.9	68.6	63.2
4: increase	10.3	16.0	18.6	17.4	18.4	26.5
5: strong increase	0.4	0.9	0.8	0.6	0.9	0.9
mean	2.66	2.95	3.04	2.98	3.07	3.17

Since the number of answers in the categories *strong decrease* and *strong increase* is quite low, we decided to combine categories 1 and 2 and categories 4 and 5. This means that we have three possible outcomes for the dependent variable EXP_t: p equals 3 in equation (2). The explanatory variables in the equation for the underlying unobserved variable include (dummies for) income changes in the past, sex, age, actual income, and dummy variables for the labor market status of the head of household and spouse. We refer to Tables B1 and B2 in Appendix B for definitions and summary statistics of these variables.

First we estimate the ordered *random effects* model described in section 2.1. We fix $m_1 = -1$ by means of normalization. The random effects α_i are assumed to be normally distributed. For the distribution of the error terms $u_{i,t}$, we choose the (standard) logistic distribution. We also estimated the random effects model with a normally distributed $u_{i,t}$. The results were almost the same. That is, the same parameters were significant and all these significant parameters had the same sign. Vuong's (1989) model selection test, however, suggests that the model with logistic $u_{i,t}$ fits the data significantly better than the model with normally distributed $u_{i,t}$. [7]

The total number of observations in the pooled sample is equal to 6845. Estimation results are presented in Table 2a. No restrictions are imposed upon the slope coefficients across the various waves. The estimates here are very similar to the estimates obtained when estimating the cross-section model for each separate wave. The only joint elements are the boundary m_2, and the variance of the random effect, which picks up about 20% of the total error variance (σ_u^2, the variance of the standard logistic distribution, is equal to $\pi^2/3$). Joint estimation has the advantage that stability of coefficients over time can be tested straightforwardly. The test results are presented in the final column of Table 2a. [8]

[7] The realization of the test statistic, that should be compared with a critical value of the standard normal distribution, is equal to 14.8.

[8] All tests are Wald tests, based upon imposing $T-1=5$ restrictions in the general model.

Table 2a : Estimation results for the ordered random effects model

DEPENDENT VARIABLE: EXP_t ($t = 84, \ldots, 89$)							
Number of Observations: 6845							
Variable	1984	1985	1986	1987	1988	1989	
CONSTANT	3.91*	4.30*	4.45*	3.81*	4.23*	3.96*	NR
DECR_1	-1.79*	-1.20*	-0.76*	-1.26*	-0.76*	-0.46*	R
INCR_1	1.41*	1.08*	1.14*	0.98*	0.98*	1.04*	NR
SEX	-0.21	-0.10	-0.30*	-0.27*	-0.25*	-0.10	NR
AGE	-1.35*	-0.98*	-0.89*	-0.75*	-0.84*	-0.75*	NR
AGE2	0.12*	0.07*	0.05*	0.04	0.05*	0.05*	NR
LOG_INC	0.30*	0.11	0.09	0.34*	0.26*	0.11	R
DUNEM	-1.00*	-1.27*	-1.03*	-0.68*	-0.73*	-0.04	R
DDIS	-1.65*	-1.30*	-1.04*	-0.98*	-0.92*	-0.05	R
DRET	-0.27	-0.34	-0.07	0.11	-0.02	-0.20	NR
DOTH	-0.07	-0.27*	-0.42*	-0.03	-0.34*	-0.36*	R
DTWO	-0.54*	-0.21	-0.26*	-0.32*	-0.18	-0.27*	NR
σ_α^2	0.75*						
m_2	3.18*						

1) * = significant at 5 % level.
2) Null hypothesis: coefficient corresponding to explanatory variable does not vary over time; R = rejected, NR = not rejected (significance level: 5%).

The 1984 estimates are similar to those in Das and Van Soest (1996). Many of these appear to remain stable over time. However, a joint test on the stability of the coefficients AGE and AGE2 rejects the null hypothesis that the age pattern remains constant over time. This suggest that there might be some cohort effect. Households with a female head tend to be less optimistic than other one earner households: the coefficient of SEX is negative and significant in three of the six years. [9] Except for 1985 and 1988, two earner households have significantly lower expectations of income changes than other households headed by males. For none of the years, retired family heads are significantly different from working heads. For the dummy variables corresponding to unemployed and disabled family heads, stability over time is rejected. Both reveal a similar tendency: unemployed

[9] For married couples, the head of household is by definition the husband.

and disabled heads are significantly more pessimistic than workers (with the same income) in the first five years, but the differences decline and have basically disappeared in the last wave. For the disabled, this may well reflect anticipation to the institutional changes in disability benefit access and levels that started in 1985 and were completed in 1987. For the unemployed, it probably reflects larger expected chances of finding a job due to the upswing of the business cycle.

Those who experienced an income decrease in the past have a larger probability of expecting another income decrease than others (ceteris paribus). This effect is not stable over time and tends to become smaller, but it remains significant throughout the time period under consideration. On the other hand, those who experienced an income increase tend to remain less pessimistic than others, and the difference with those whose income did not change during the last twelve months (the reference group) remains stable over time.

Stability over time of the relation between income expectations and the level of actual income LOG_INC (objectively measured), is rejected at the 5 % level. Still, the effect is always positive, and significant in three out of the six years. This suggests a tendency of increasing income inequality: the rich relatively more often expect to get richer, the poor expect to get poorer. We come back to this below, where we link this to the findings for the fixed effects model.

In the fixed effects specification, the assumption of independence between the individual effect and the covariates are relaxed (see section 2.2). We normalized the constant term and the coefficients of the variables SEX, AGE, and AGE2, which do not vary over time or vary over time in a deterministic way, to zero for the first wave. Using Wald tests for each of these variables separately, we found that these variables were insignificant at the 5 % level for the other waves. The results we present are those obtained after excluding these variables. Note that with the estimates of the fixed effects specification we do not use data on the households that provided all information in just one wave.

In our application the number of categories p is equal to 3: decrease ($\mathrm{EXP}_t < 3$), no change ($\mathrm{EXP}_t = 3$), and increase ($\mathrm{EXP}_t > 3$). As mentioned in section 2.2 we summarize the ordered categories into two categories so that we can use the conditional logit procedure. This means that there are two possible summaries: *2* versus *3* and *4*, and *2* together with *3* versus *4*. By using a minimum distance step we combine these two estimators to get the final estimates for the β_t's. These final results are shown in Table 2b.

Table 2b : Estimation results for the ordered fixed effects model

DEPENDENT VARIABLE: EXP_t ($t = 84, \ldots, 89$)							
Number of Observations: 5185							
Variable	1984	1985	1986	1987	1988	1989	
DECR_1	-1.62*	-0.71*	-0.36*	-0.89*	-0.45*	-0.35*	R
INCR_1	0.60*	0.48*	0.63*	0.31*	0.39*	0.55*	NR
LOG_INC	-0.57*	-0.17*	-0.15*	-0.01	-0.02	-0.04	R
DUNEM	-0.76*	-0.42	-0.58*	-0.30	0.14	0.78*	R
DDIS	-1.66*	-0.74*	-0.55*	-0.73*	-0.33	0.88*	R
DRET	0.22	0.07	0.46*	-2E-3	0.48*	1.13*	R
DOTH	-0.12	-0.06	-0.23*	0.15	-0.06	0.84*	R
DTWO	-0.39*	-0.31*	-0.26*	-0.48*	-0.27	-0.27*	NR

1) * = significant at 5 % level.
2) Null hypothesis: coefficient corresponding to explanatory variable does not vary over time; R = rejected, NR = not rejected (significance level: 5%).

For the variables referring to realized income changes in the past, the results are basically the same as those for the random effects model. Those whose income decreased in the past are significantly more pessimistic, and those whose income increased are more optimistic than those whose income remained unchanged. The results for the labor market status variables are also similar to those in Table 2a. The only remarkable difference is found for $t = 89$. In Table 2a DUNEM, DDIS, and DRET are not significantly different from zero while in Table 2b all these parameters are significantly positive. This suggests that in 1989 those heads of households who became unemployed, disabled or retired are less pessimistic about future income growth than the employed heads.

Only for the variable LOG_INC we find a result which is substantially different from that in the random effects model. The coefficient is negative instead of positive, and significant in three out of the six waves. An explanation is that the fixed individual effect is positively correlated with income. Thus those with higher 'permanent' incomes are on average more optimistic than others. This is revealed by the positive sign in the random effects model. It suggests that heads of households expect that differentials in incomes per year between those with high and those with low permanent income tend to increase over the life cycle. The estimates of the fixed effects model then tell us that, conditional on the fixed effect and permanent income, those whose income is unusually low in a given period often expect an income rise while those with relatively high income expect their income to fall. This corresponds to the notion that the deviation between actual income and permanent income can be seen as transitory, and that the expected change in transitory income is negatively related to the level of transitory income.

The fixed effects specification is a generalization of the random effects model. The two can be compared using a Hausman test. If the random effects model is correctly specified, the random effects ML estimates for the β_t are consistent and asymptotically efficient. The estimates of the fixed effects model are consistent as long as the fixed effects specification is correct. The Hausman test is based upon the differences of the two sets of parameter estimates. The test leads to a clear rejection of the random effects specification, on every sensible significance level. [10]

[10] We compared the random effects model and the fixed effects model allowing for time variation of all the parameters. Details of the test are available upon request from the authors.

4 Comparing expectations with realizations

Family heads were also asked to answer the question

> *Did your household's income increase, decrease, or remain unchanged during the past twelve months?*

The possible answers, which we denote by PREV_t ($t = 84, \ldots, 89$), are the same as for EXP_t. The distribution of the answers is given in Table 3.

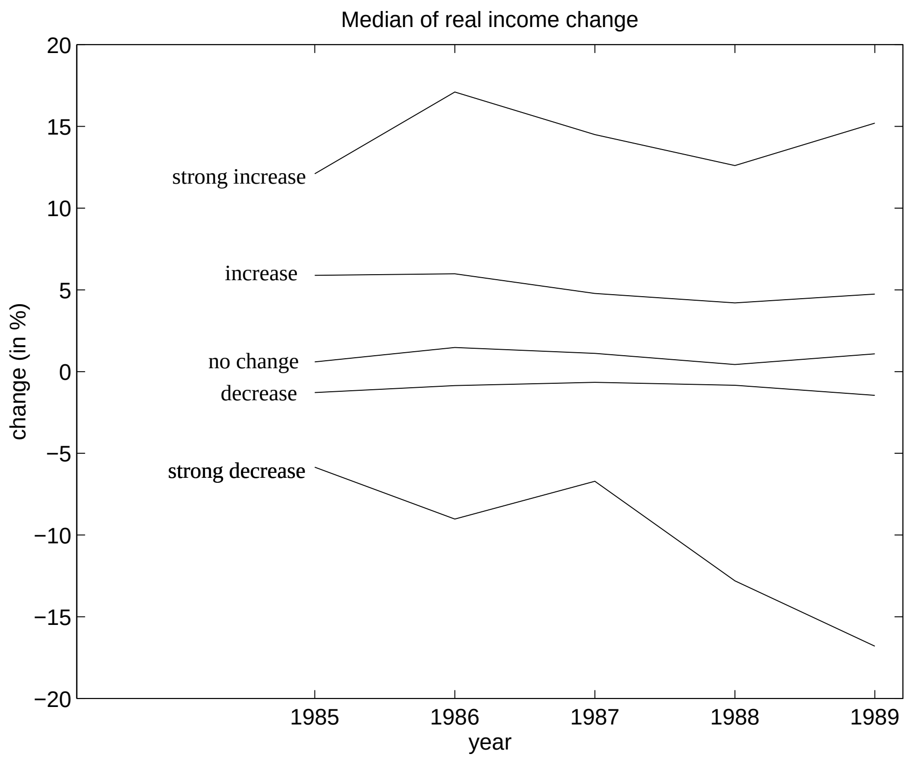
Table 3 : Univariate frequencies (in %) of PREV_t ($t = 84, \ldots, 89$)

PREV_t	84	85	86	87	88	89
1: strong decrease	11.7	9.1	4.9	5.5	4.4	3.8
2: decrease	24.6	16.9	10.7	15.2	9.1	6.9
3: no change	51.6	53.9	56.3	55.8	60.2	56.1
4: increase	9.0	15.7	23.1	19.0	20.4	26.1
5: strong increase	3.1	4.3	5.0	4.5	5.9	7.0
mean	2.67	2.89	3.12	3.02	3.14	3.26

If we compare Table 3 with Table 1 we see that the dispersion in realized income changes is much larger than in expected income changes. There are quite a lot of households who experienced a strong decrease or a strong increase. This is not surprising, since the expected income change refers to some location measure of the household's (subjective) income change distribution, while the realization is one draw from the (actual) distribution of income change. The dispersion in the latter is therefore not only due to variation in income growth distributions across families, but also to the uncertainty of the income change for a given household. Further, we see that the mean values for PREV_t follow the same increasing pattern as the mean values for EXP_t, with a dip in 1987. The relatively low (mean) expectation in 1987 (for the twelve months after October 1987) might thus be explained by the dip in realized income changes for the twelve months preceding October 1987.

Figure 1 shows the relation between the answers to the subjective income change question and the objectively measured change in actual real total family income over the same time period (using the consumer price index for each year). We present the median real income change for families with given value of PREV.

Figure 1 : Relation between the answers to the subjective income change question and the objectively measured change in actual real total family income.



The results are quite stable over time, except for those who experienced a large decrease. For those who reported no change (PREV = 3), the median real income change varies between 0.4% and 1.5%. For those who reported an income decrease, the medium real change varies from -1.5% to -0.5%; for those who reported an increase, it varies from 4.2% to 6.0%. These numbers are more stable if we look at real income changes rather than if we would consider nominal income changes. In Das and Van Soest (1996) we already argued that the subjective answers reflect real rather than nominal changes. Figure 1 provides further evidence to support this conclusion. For those reporting a strong increase, the median varies between 12.1% and 17.1%. Only for those who reported a strong decrease, the pattern seems nonstationary, and the median falls from -5.9% to -16.8%. Note, however, that this group has become quite small in 1989 (see Table 3).

Although the questions are not very well specified, it seems reasonable to assume that the head of household has the same concept in mind while answering the questions on PREV_t and EXP_t. Due to the panel nature of the data set we can compare the expectation of income change (provided in wave t-1) with the realization for the same time period (provided in wave t). If PREV_t is larger than EXP_{t-1} then we can say that the head of household *ex post* appears to have underestimated household income growth. Analogously, if PREV_t is smaller than EXP_{t-1} then the income growth is overestimated.

Table 4 : Frequencies (in %) of under- and overestimating future income changes

	underestimation	overestimation	Test-statistic
1984-1985	34.9	15.4	12.8
1985-1986	29.3	15.9	9.9
1986-1987	22.5	21.5	0.9
1987-1988	29.2	14.6	13.1
1988-1989	28.9	12.5	15.6

Note: A conditional sign test is carried out to test whether the probability of overestimating equals the probability of underestimating future income growth. The third column displays the test-statistic that should be compared with critical values from the standard normal distribution.

Table 4 shows the frequencies of households who under- and overestimated their income changes. In all cases, we see that the percentage of families underestimating exceeds the percentage of families overestimating future income growth. Except for 1986-1987, this difference is highly significant. We find it hard to believe that unanticipated macro-economic shocks explain the fact that this happens several times in a row. Although macro-economic changes may well be correlated over time, we see no reason why the unanticipated element in them should.

A possible weakness of the confrontation of expectations with realizations given above might be implied by the vague wording of the question. If someone has experienced strong decreases in the past, one may have got used to it, and won't use the word *strong* again (habit formation effect). To eliminate this problem, we recalculated the test-statistics in Table 4, but now after combining the categories 1 and 2 and the categories 4 and 5, so that the difference between *strong* and *moderate* is eliminated. The values of the test-statistics for the five years are then given by $14.2, 10.3, 0.1, 12.3$, and 14.8. Again, the underestimation is significant in four years. Only for 1986-1987 the result is not significant.

In Table 5 we present the estimates of an ordered response panel data model with fixed effects explaining the deviation $\mathrm{DEVIATION}_t = \mathrm{EXP}_{t-1} - \mathrm{PREV}_t$ between income change expectation and income change realization for the same time period. The model and estimation strategy are those discussed in section 2. The possible values of the dependent variable range from -4 (strong underestimation of future income) to 4 (strong overestimation). This would lead to 8 possible conditional logit estimates. However, because of the low numbers of observations in the extreme categories and for computational convenience, we only used two summaries of the data: $\mathrm{DEVIATION}_t < 0$ versus $\mathrm{DEVIATION}_t \geq 0$ and $\mathrm{DEVIATION}_t \leq 0$ versus $\mathrm{DEVIATION}_t > 0$. The two conditional binary logit estimates are combined using minimum distance.

Again, for each variable, a Wald test is performed on stability over time of the corresponding parameter. Moreover, an additional Wald test is carried out to test whether all parameters corresponding to a specific explanatory variable are equal to zero. Except for the variables LOG_INC, DUNEM, and DTWO both tests reject the null hypothesis. The unemployed heads do not significantly differ from working heads and heads of two earner households do not underestimate more or less than other male family heads. Disabled heads have tended to underestimate significantly more than employed heads in 1988 and 1989. An interpretation of this is that people expected stronger consequences of the reforms of the system of disability benefits.

Table 5 : Estimation results for the ordered fixed effects model

DEPENDENT VARIABLE: DEVIATION$_t$ = EXP$_{t-1}$− PREV$_t$ ($t = 85, \ldots, 89$)							
Number of Observations: 4243							
Variable	1985	1986	1987	1988	1989	$H_0^{(1)}$	$H_0^{(2)}$
DECR_1	-0.60*	-0.97*	-0.78*	-1.16*	-1.09*	R	R
INCR_1	1E-3	0.33*	0.73*	0.87*	0.70*	R	R
LOG_INC	0.18*	0.14	0.06	0.11	0.02	NR	NR
DUNEM	-0.26	0.06	0.32	-0.12	-0.37	NR	NR
DDIS	0.07	-0.20	0.15	-0.70*	-0.82*	R	R
DRET	-0.42	-0.24	0.85*	-0.05	-0.08	R	R
DOTH	-0.49*	-0.08	0.12	-0.03	-0.31*	R	R
DTWO	4E-3	0.34*	0.12	0.03	0.22	NR	NR

1) * = significant at 5 %.
2) Hypothesis $H_0^{(1)}$: coefficients corresponding to explanatory variable do not vary over time; Hypothesis $H_0^{(2)}$: all the coefficients corresponding to explanatory variable are equal to 0 (R = rejected, NR = not rejected, significance level = 0.05).

The effects of DECR_1 and INCR_1, the variables indicating an income decrease or increase in the past, are not stable over time. [11] Still, the effect of DECR_1 is significantly negative and the effect of INCR_1 is significantly positive in all years. This implies that those whose income has fallen have a larger probability of underestimating than others. This result was also found by Das and Van Soest (1996). We find that this result is robust over time.

[11] No account has been taken of potential endogeneity of these variables.

The main findings of this analysis are the following. First, the number of people underestimating future income growth is larger than the number of people overestimating income growth. Second, the tendency to underestimate varies with labor market status and income change history. In particular, those whose income has fallen in the past tend to underestimate future income growth. Various explanations could be given for this finding. First, it could be a statistical artifact due to the fact that we are comparing an *ex ante* location measure with an *ex post* realization. Even if households' subjective and actual income change distribution are the same, some heads of households will overestimate and some will underestimate, and, due to the categorical nature of the data, the numbers of those who underestimate and overestimate are not necessarily the same (see Manski, 1990, p. 937). Although this might explain why we find an overall tendency of underestimation, we do not think that this argument can explain why particularly those whose income fell in the past underestimate.

The second explanation would be the existence of (unexpected) shocks which are correlated across households with certain characteristics. For example, if macro-economic growth rates are larger than expected for various years in a row, this could explain why we find underestimation on average. Again however, it seems hard to imagine that positive shocks are particularly relevant for those whose income has fallen in the past.

The third explanation is that people's expectations are not rational, and that households whose income has fallen are simply too pessimistic. This could mean that heads of household tend to view negative income changes too much as permanent.

5 Conclusions

We have analyzed subjective data on income change expectations and realizations using panel data covering the period 1984 – 1989. Comparing the subjective data with information on actual income suggests that, on average, the data are consistent with the notion that people consider percentage changes in real income. For all panel waves, we find that income growth expectations are strongly affected by previous income changes. The impact of labor market status variables is less stable over time, and this can partly be explained by institutional changes in the time period considered. Comparing random effects and fixed effects estimates of the coefficient of the actual income level leads to the conclusion that those with higher permanent incomes generally have higher expected income growth than others. On the other hand, those with low or negative transitory income often expect an income rise, while those with high transitory income expect their income to fall.

Comparing expected and realized income changes for the same time period, we find for all waves but one that on average, future income growth was significantly underestimated. In particular, people whose income decreased in the recent past tend to be too pessimistic. It seems hard to imagine that this is caused by unanticipated macro-economic shocks. First, we cannot think of shocks which would affect those with a specific income change (and not a specific income level). Second, the effect is remarkably persistent over time. A plausible alternative explanation seems to be that people's expectations are not rational, and that negative transitory incomes are too often considered to be permanent.

Our results thus cast doubt on using the assumption of rational expectations, a common assumption in many empirical studies of life cycle models. Moreover, our results suggest that subjective survey questions contain valuable additional information, which can be used to replace this assumption. Incorporating this in a life cycle model thus seems a promising topic of future research.

Appendix A

In this appendix we present some details on the estimation procedure in the ordered response panel data model with fixed individual effects. For details on the binary case we refer to Chamberlain (1980).

First we combine adjacent categories so that the dependent variable $y_{i,t}$ is summarized as a binary variable. There are $p-1$ of such combinations and for each combination we use the conditional logit method proposed by Chamberlain (1980). Under some regularity conditions, all the conditional ML estimators for the parameter vector of interest $\beta \in \mathbb{R}^{kT}$ are consistent and asymptotically normal:

$$\sqrt{n}(\hat{\beta}_s - \beta) \rightarrow \mathrm{N}(0, (\mathrm{E}\{l_s l_s'\})^{-1}), \qquad s = 1, \ldots, p-1,$$

where l_s is the score vector corresponding to combination s. The fixed effects estimator of β is then obtained by a minimum distance step:

$$\hat{\beta}_{FIX} = \mathrm{argmin}_{\beta} \frac{1}{2} [\begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{pmatrix} - \begin{pmatrix} \beta \\ \vdots \\ \beta \end{pmatrix}]' W^{-1} [\begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{pmatrix} - \begin{pmatrix} \beta \\ \vdots \\ \beta \end{pmatrix}].$$

The optimal weighting matrix W is given by $\Omega = [\omega_{a,b}]$ where

$$\omega_{a,b} = (\mathrm{E}\{l_a l_a'\})^{-1} \mathrm{E}\{l_a l_b'\} (\mathrm{E}\{l_b l_b'\})^{-1}, \qquad a, b = 1, \ldots, p-1.$$

In the actual calculations we replace the expectations by their sample analog and the true parameter values by their estimations. Since $\hat{\beta}_{FIX}$ is a linear combination of the consistent estimators $\hat{\beta}_1, \ldots, \hat{\beta}_{p-1}$, the fixed effects estimator is consistent. The asymptotic distribution of the fixed effects estimator is given by

$$\sqrt{n}(\hat{\beta}_{FIX} - \beta) \rightarrow \mathrm{N}(0, (D'\Omega^{-1}D)^{-1}),$$

where

$$D = \begin{bmatrix} I_{Tk \times Tk} \\ \vdots \\ I_{Tk \times Tk} \end{bmatrix} \in \mathbb{R}^{(p-1)Tk \times Tk}.$$

Appendix B

Table B1: reference list variables.

EXP_t	Answer to the question : "What will happen to your household's income in the next twelve months ?" Possible answers are: strong decrease (1); decrease (2); no change (3); increase (4); strong increase (5). The subindex t runs from 84 till 89 (where 84 corresponds to the year 1984, etc.).
$PREV_t$	Answer to the question : "Did your household's income increase, decrease, or remain unchanged during the past twelve months ?" Possible answers: see EXP_t.
DECR_1	Dummy variable related to $PREV_t$: DECR_1 = 1 if $PREV_t$ is equal to 1 or 2; 0 otherwise.
INCR_1	Dummy variable related to $PREV_t$: INCR_1 = 1 if $PREV_t$ is equal to 4 or 5; 0 otherwise.
SEX	Sex head of household: 1 = male; 2 = female. If husband and wife are present, the husband is by definition head of household.
AGE	Age head of household in tens of years.
LOG_INC	Natural logarithm of net household income where net household income is in tens of thousands (per year). The survey contains accurate information on income from about twenty potential sources for each individual. After tax household income was constructed by adding up all individual income components of all family members and some specific household components (such as child allowances).

Dummy-variables corresponding to labor market status head of household:

DEMP	1 if head of household is employed; 0 otherwise.
DUNEM	1 if head of household is unemployed; 0 otherwise.
DDIS	1 if head of household is disabled; 0 otherwise.
DRET	1 if head of household is retired; 0 otherwise.
DOTH	DOTH=1-DEMP-DUNEM-DDIS-DRET

Dummy-variable corresponding to labor market status of spouse:

DTWO	1 if household is a two-earner household; 0 otherwise.

Table B2: Mean values (standard deviations in parentheses).

t	84	85	86	87	88	89
Nr. Obs	2683	2787	3850	3899	4059	4133
EXP_t	2.66	2.95	3.04	2.98	3.07	3.17
	(0.76)	(0.68)	(0.64)	(0.67)	(0.61)	(0.64)
$PREV_t$	2.67	2.89	3.13	3.02	3.14	3.26
	(0.90)	(0.92)	(0.85)	(0.86)	(0.83)	(0.83)
DECR_1	0.36	0.26	0.16	0.21	0.14	0.11
INCR_1	0.12	0.20	0.28	0.24	0.26	0.33
SEX	1.20	1.19	1.19	1.23	1.23	1.23
Age head of household	46.6	46.1	45.6	47.1	47.0	46.9
	(17.0)	(16.4)	(16.2)	(17.0)	(16.9)	(17.0)
Net household income	3.48	3.57	3.64	3.79	3.71	3.79
(in Dfl. 10,000)	(1.98)	(2.24)	(2.12)	(2.98)	(2.32)	(2.21)
DEMP	0.554	0.545	0.587	0.528	0.543	0.575
DUNEM	0.045	0.037	0.030	0.030	0.022	0.026
DDIS	0.068	0.075	0.063	0.069	0.061	0.063
DRET	0.158	0.143	0.183	0.230	0.229	0.193
DTWO	0.204	0.216	0.253	0.230	0.235	0.245

References

Alessie, R. and A. Lusardi (1996), Saving and Income Smoothing: Evidence from Panel Data, *European Economic Review*, forthcoming.

Alessie, R., A. Lusardi, and T. Aldershof (1996), Income and Wealth over the Life Cycle: Evidence from Panel Data, *Review of Income and Wealth*, forthcoming.

Butler, J.S. and R. Moffitt (1982), A Computationally Efficient Quadrature Procedure for the One-factor Multinomial Probit Model, *Econometrica*, 50, pp. 761-764.

Carroll, C.D. (1994), How does future income affect current consumption?, *Quarterly Journal of Economics*, 109, pp. 111-147.

CBS (1991), *Sociaal-economisch panel onderzoek*, Centraal Bureau voor de Statistiek, Voorburg.

Chamberlain, G. (1980), Analysis of Covariance with Qualitative Data, *Review of Economic Studies*, 47, pp. 225-238.

Chamberlain, G. (1984), Panel Data, in *Handbook of Econometrics* (Vol. II.), eds. Z. Griliches and M.D. Intiligator, Amsterdam: North-Holland, pp. 1247-1318.

Crouchley, R. (1995), A Random-Effects Model for Ordered Categorical Data, *Journal of the American Statistical Association*, 90, pp. 489-498.

Das, M. and A. van Soest (1996), Expected and Realized Income Changes: Evidence from the Dutch Socio-Economic Panel, *Journal of Economic Behavior and Organization*, forthcoming.

Deaton, A. (1992), *Understanding Consumption*, Oxford University Press, Oxford.

Dominitz, J. and C.F. Manski (1996), Using Expectations Data to Study Subjective Income Expectations, Department of Economics, University of Wisconsin-Madison.

Gibbons, R.D., D. Hedeker, S.C. Charles, and P. Frisch (1994), A Random-Effects Probit Model for Predicting Medical Malpractice Claims, *Journal of the American Statistical Association*, 89, pp. 760-767.

Guiso, L., T. Jappelli, and D. Terlizzese (1992), Earnings uncertainty and precautionary saving, *Journal of Monetary Economics*, 30, pp. 307-337.

Guiso, L., T. Jappelli, and D. Terlizzese (1996), Income Risk, Borrowing Constraints and Portfolio Choice, *American Economic Review*, forthcoming.

Lusardi, A. (1993), Precautionary Saving and Subjective Earnings Variance, VSB Progress Report 16, CentER for Economic Research, Tilburg University.

Manski, C.F. (1990), The use of intentions data to predict behavior: a best-case analysis, *Journal of the American Statistical Association*, 85, pp. 934-940.

Neyman, J. and E.L. Scott (1948), Consistent Estimates Based on Partially Consistent Observations, *Econometrica*, 16, pp. 1-32.

Pfeiffer, F. and W. Pohlmeier (1992), Income, Uncertainty and the Probability of Self-Employment, *Recherches Economiques de Louvain*, 58, pp. 265-281.

Vuong, Q.H. (1989), Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, 57, pp. 307-333.